

College & Research Libraries. 2009, vol. 70, no. 6. (November): 568-577.

Print ISSN: 0010-0870 Online ISSN: 2150-6701

<http://crl.acrl.org/>

<http://crl.acrl.org/content/70/6.toc>

<http://crl.acrl.org/content/70/6/568.full.pdf+html>

©2009 Sarah Anne Murphy, Sherry Engle Moeller, Jessica R. Page, Judith Cerqua, and Mark Boarman

Leveraging Measurement System Analysis (MSA) to Improve Library Assessment: The Attribute Gage R&R

Sarah Anne Murphy, Sherry Engle Moeller, Jessica R. Page, Judith Cerqua, and Mark Boarman

Measurement System Analysis (MSA) provides decision makers with a useful suite of tools for understanding whether variation should be attributed to an assessment system itself or the actual item or program being assessed. This paper introduces the Attribute Gage R&R, using a study of The Ohio State University Libraries' mechanism for measuring quality in e-mail reference transactions as an example. An ideal tool for examining assessment programs that require subjective interpretation, the Attribute Gage R&R can assist library organizations in understanding their processes and validating the utility of data collected through their measurement systems.

Libraries utilize measurement systems for a multitude of purposes: to collect and analyze data, such as the number of directional versus reference questions, for reporting purposes; to understand patrons' use of technology, such as the library's Web site, to improve its design and usability; and to improve and control both internal and public transactional processes, in areas such as cataloging and circulation. Gathering reliable data, when processes require subjective inspection or validation, is a major challenge for all librarians, especially those who interact directly with the public. This paper focuses on Measurement System Analysis (MSA), specifically the Attribute Gage R&R, as a useful suite of tools for understanding variation introduced by the assessment system itself and not the actual item or program being assessed. It is a continuation of an earlier exploration of the application of Lean Six Sigma, a business improvement philosophy and methodology, in the academic library environment, as a means to nurture and sustain a culture of assessment and change.¹

The previous project utilized the *RUSA Guidelines for the Behavioral Performance of Reference and Information Service Providers* (RUSA Guidelines) as operational definitions for measuring quality in e-mail reference transactions.² The experience revealed a wide variety of interpretations of these guidelines by librarians and paraprofessional assistants at The Ohio State University. Following Lean Six Sigma principles, a team was established to draft local standards for interpreting and applying the RUSA Guidelines. The paper will address the team's experience and begin with a brief literature review to introduce MSA, what it is, and why it is important. It will continue by describing the methods to conduct an Attribute Gage R&R analysis for a transactional service process and the tool's utility for establishing shared operational definitions. Results of the analysis will then be provided, followed by a discussion highlighting the lessons learned.

Literature Review

For years, librarians have struggled to agree on a shared definition for a reference question and, further, what constitutes a quality answer to a reference question.³ The RUSA Guidelines recognize that the success of a reference transaction cannot only be measured by the quality of the

information provided, but the positive behavior of the individual with whom the patron interacted.⁴ The Guidelines have been applied to a number of reference evaluation studies, in face-to-face, chat, and e-mail reference environments.⁵

While “what gets measured gets done” is a popular business mantra, an experienced practitioner of Lean Six Sigma will ask, “Did you check your measurement instrument?” This is because a measurement system contaminated with error will understandably result in flawed decisions, potentially affecting what does or does not get done.⁶ Traditionally used in manufacturing, MSA uses tools such as the Attribute Gage R&R to mathematically model the capability of a measurement system. Specifically, the purpose of such analysis is to determine how much observed variability may be attributed to the measurement system itself, as opposed to the process being monitored.

For processes such as the reference transaction, which require the subjective interpretation and application of guide-lines to evaluate, there is a “potential for variability among inspectors and even variability by the same inspector over a period of time.”⁷ The ‘Rs’ in Attribute Gage R&R refer to repeatability and reproducibility. Thus, one individual evaluating the quality of a reference transaction should be able to replicate the same results if evaluating the same reference transaction at a later date and time. This is defined as repeatability, or intra-appraiser agreement. A colleague evaluating the same reference transaction, using the same criteria, should be able to record the same result. This is referred to as reproducibility, or inter-appraiser agreement. Attribute refers to the type of Gage R&R used for analyzing systems using discrete or categorical data. Thus, if the librarian correctly identified a question as directional, the appraiser would record “yes.” If the question should have been identified as a reference question, the appraiser would record “no.” The purpose of conducting the Attribute Gage R&R is to determine whether variation within the measurement system can be attributed to the appraisers, the measurement instrument, or the evaluation process itself. If operational definitions supporting the measurement instrument, for example, are ambiguous or not clearly understood and consistently applied by appraisers, low repeatability and reproducibility scores will result.

Methods

Like the previous project, which sought to improve turnaround time and the quality and consistency of communications for the OSU Libraries’ e-mail reference service, this project was also organized using Lean Six Sigma principles and tools.⁸ The project mission was to create local standards for interpreting and applying the RUSA Guidelines that could be incorporated into the policies and procedures for the OSU Libraries’ e-mail reference service. To accomplish this mission, the Libraries’ mechanism for classifying question type and evaluating question conformance with the RUSA Guidelines for Approachability, Listening/Inquiring, and Follow-up was examined. The previous project used these definitions to monitor both the appropriate distribution of questions throughout the OSU Libraries system and the quality of answers.

The project team consisted of two librarians, two paraprofessional staff members, and the Coordinator of Research and Reference. The OSU Libraries uses OCLC’s QuestionPoint service to centrally manage questions received through its Ask-A-Question Web site. A paraprofessional is responsible for monitoring all incoming questions, answering the directional and basic reference questions directly, and referring questions requiring in-depth knowledge to subject librarians to answer. Typically, a sample of twenty to thirty questions is required to conduct an Attribute Gage R&R; however, to maintain a representative sample of answers, a larger sample size was required

for this analysis. For the first stage of the project, 100 of the 586 questions received during winter quarter from January 1, 2008, to March 17, 2008, were examined.

When an Attribute Gage R&R is conducted, there is an option to examine both inter-appraiser reproducibility and intra-appraiser repeatability alone, and in relation to a standard, or expert, appraiser. For this project, the Coordinator for Research and Reference, as the individual responsible for establishing policy for the OSU Libraries' reference services, served as the standard. A meeting was convened to review the RUSA Guidelines for Approachability, Listening/Inquiring, and Follow-Up behaviors, along with the OSU Libraries definitions for directional, basic, and specialist reference questions that were based on ARL definitions and used in the previous project. Markers for what constituted an incomplete/incorrect answer and service denial were also reviewed. The project team then reviewed the sample, using a Microsoft Excel spreadsheet to assign a question to the directional, basic, or specialist reference category, and mark whether the answer conformed to the RUSA Guidelines for Approachability, Listening/Inquiring, and Follow-up, if the answer was incorrect/ incomplete, or if the patron was inappropriately denied service.

Typically in a Gage R&R, all inspectors individually review the same sample twice. It quickly became apparent, as data came in during the first round that inter-appraiser reproducibility for the coding categories ranged from 17 percent to 89 percent. (See table 1.) Because of the three to four hours required to read and code the questions and answers, the project leader halted the analysis before the second review. A second meeting was convened to review coding disparities and discuss team members' reasoning for coding certain questions as conforming or not conforming, and others as basic reference instead of directional or specialist. As a result of this discussion, the team decided it could not agree on a uniform interpretation of the RUSA Guidelines and what constituted an incomplete answer. The team also disagreed on the interpretation of the definitions for coding questions as directional, basic, or reference. Some team members, for example, believed all questions related to using the library catalog should be coded as basic reference, while others thought this category should be reserved for questions requiring basic author, title, and subject searches. Other team members coded questions involving the Libraries' EZ-proxy software and off-campus authentication as directional, while their colleagues marked these as basic reference.

Through the process of reviewing the coding disparities, clarifying language was added to the definitions for directional, basic, and specialist reference along with concrete examples. Questions such as "How do I place a hold on a book?" were determined to be basic reference, leading the team to insert the text "Questions that require instruction on using the library catalog, such as how do I recall a book, should be coded as basic reference" into the basic reference definition. Coding disparities highlighted by team members' disagreement over whether to code certain questions as directional or basic reference were resolved by adding examples of information contacts to code as directional to the directional question definition.

The team also decided to simplify the coding instrument, by consolidating the RUSA categories with the incomplete/incorrect answer category. A new definition for what constituted a quality answer in the opinion of the project team was created, with concrete examples. The RUSA Follow-up category, however, was maintained separately, at the Coordinator of Research and Reference's request, as the previous project indicated that failure to include follow-up language was the most frequently occurring problem with question answers. The service denial category was also retained.

The analysis was then run again, using the same question sample. The project leader again

halted the analysis early, as preliminary data indicated inter-appraiser reproducibility for the new categories ranged from 34 percent to 76 percent. (See table 1.) The project team reconvened a third time to review coding disparities, the definitions, and each individual's interpretation of the definitions. During this session, additional clarification language was added to the definitions and standards. Coding fatigue was also discussed, as the team members noted they would be reading the same questions and answers for the third time. It was decided to try the analysis one last time, using a new sample. Of the 474 questions and answers received during summer quarter, from June 16, 2008, to August 29, 2008, another 100 were randomly selected for analysis. It was decided that the full Attribute Gage R&R would be conducted during this round, regardless of the preliminary results. On completion, the data gathered in the Microsoft Excel spreadsheets were then transferred into Minitab for analysis using the Attribute Agreement Analysis module.

TABLE 1 Percentage of Preliminary Inter-Appraiser Agreement (Reproducibility) for 1st and 2nd Rounds of Attribute Gage R&R		
Variable	Round 1	Round 2
Directional	64	34
Basic Reference	17	38
Specialist Reference	33	53
Quality of Answer	n/a	46
Incorrect/Incomplete Answer	55	n/a
Incorrect Referral	79	n/a
RUSA: Approachability	64	n/a
RUSA: Listening/Inquiring	83	n/a
RUSA: Follow-up	37	35
Service Denial	89	76

Results

Most Lean Six Sigma practitioners note that acceptability levels for any Gage R&R varies depending on the situation, but generally the target for repeatability and reproducibility is 80 percent to 90 percent.⁹ "The most important part of any such exercise," reminds Samuel E. Windsor, "is to turn the raw data into either a validation of the system or an action plan to fix the system."¹⁰ Results for intra-appraiser repeatability and inter-appraiser agreement for the third and final attempt for the Attribute Gage R&R are listed in table 2. Individual appraisers assigned questions to the same directional, basic reference, and specialist reference categories with 87 percent to 100 percent consistency. Between-appraiser agreement for these categories ranged from 69 percent to 88 percent. While further work for fine-tuning of the operational definitions for classifying question type is required, this is a significant result, considering the preliminary reproducibility scores during the first attempt for the Gage R&R analysis indicated 17 percent inter-appraiser agreement for assigning a question to the basic reference category and 33 percent agreement for assigning a question to the specialist reference category. (See table 1.)

TABLE 2 Percentage of Intra-Appraiser Agreement (Repeatability) and Inter-Appraiser Agreement (Reproducibility)					
	Repeatability				Reproducibility
Variable	Appraiser 1	Appraiser 2	Appraiser 3	Appraiser 4	(Between Appraiser Agreement)
Directional	87	93	99	92	70
Basic Reference	90	92	99	92	69
Specialist Reference	97	98	100	92	88
Quality of Answer	92	90	100	82	56
Follow-up	81	96	99	90	66
Service Denial	99	100	100	96	93

Between trials, appraisers were also able to consistently rate whether an answer did or did not conform to the team's definition for what constituted a quality answer, the RUSA Guidelines for follow-up behaviors, and the criteria for service denial with over 90 percent consistency in all but two instances. Between-appraiser agreement for quality of answer and RUSA follow-up, however, were 56 percent and 66 percent respectively, indicating the team is still struggling to achieve a shared understanding when applying the operational definitions for these categories. Still, progress is slowly being made in the follow-up category, as preliminary results for inter-appraiser agreement started at 37 percent in the first attempt for the Attribute Gage R&R and dipped to 35 percent in the second attempt.

Results highlighting the individual appraisers' agreement with the standard are listed in table 3. While the Coordinator of Research and Reference served as the expert appraiser, the percentage of agreement between the individual appraisers and all of the appraisers and the standard are significantly lower. For the directional, basic reference, and specialist reference classifications, appraisers individually agreed with the standard close to 80 percent of the time. Collectively, agreement between all appraisers and the standard ranged from 63 percent to 87 percent. Agreement with the expert appraiser was significantly lower for the quality of answer and follow-up categories, with individual appraisers agreeing with the standard 51 percent to 74 percent of the time, and collectively 42 percent to 45 percent. While this is a curious result, it may indicate that the Coordinator of Research and Reference more discriminately applied the written operational definitions when appraising the quality of answers and follow-up language.

TABLE 3 Percentage of Intra-Appraiser and Inter-Appraiser Agreement with Standard					
	Repeatability				Reproducibility
Variable	Appraiser 1	Appraiser 2	Appraiser 3	Appraiser 4	(Between Appraiser Agreement)
Directional	79	80	79	78	65
Basic Reference	80	75	76	77	63
Specialist Reference	93	92	95	89	87
Quality of Answer	64	67	74	60	45
Follow-up	51	56	61	51	42
Service Denial	96	100	99	94	93

Discussion/Conclusion

While work remains to refine the local operational definitions and policies for the OSU Libraries' e-mail reference service, particularly for what constitutes a quality answer, the results of the Attribute Gage R&R illustrate that the team did make progress toward developing a shared understanding. A draft Quality Standards for Virtual Reference policy resulted from the project, along with an appreciation for the need to provide consistent customer service in all reference mediums. (See Appendix A.) During the project closing review, team members noted they found critical analysis of answers to patron questions to be a valuable exercise. An unanticipated benefit of participation in the Attribute Gage R&R exercise was the opportunity to see how other librarians and paraprofessional staff handled answering questions. Team members revealed they had already started to incorporate some of the language and concepts used by colleagues into their own practice. Further, seeing good, quality answers to questions made them much more conscious of what constituted a poor answer, helping them to improve their service to customers.

The Attribute Gage R&R is an ideal tool for establishing and monitoring measurement systems, especially for processes that require subjective analysis to evaluate. It is not unusual for an organization to have to conduct a Gage R&R once, twice, or (in our case) three times to reach agreement among appraisers. The team's failure to achieve inter-appraiser agreement during the first two rounds of the analysis yielded valuable information, which was used to refine the operational definitions on which the measurement instrument was based. While perfect agreement among individual appraisers is an admirable goal, it is not realistic. Having a process to review a measurement system and give individual appraisers the opportunity to voice why they coded something as acceptable or unacceptable, however, does help move an organization toward uniform agreement.

Currently, over 40 librarians and paraprofessional staff participate in the OSU Libraries' process for receiving and answering e-mail questions. To avoid circular referrals, miscommunication, and misinforming patrons of library programs and services, quality standards can help to ensure that patrons receive a consistent library service or product. The presence of standards doesn't mean that librarians must relinquish their professional judgment in answering questions, nor does it mean librarians and paraprofessional staff must adhere to a strict institutional script when interacting with patrons. Local documentation for interpreting and applying professional guidelines, however, is useful for benchmarking and improving library service.

The team recommended that the results of the Attribute Gage R&R be shared with all librarians and paraprofessional staff receiving and answering e-mail questions and that examples of best practices for answering questions be distributed. Training and additional documentation for using the QuestionPoint system, in relation to the quality standards and best practices, is also needed. Further, the team recommended that a brief cheat sheet be created for individuals who answer a handful of questions received through QuestionPoint a year, to prompt them to incorporate elements of the standards into their replies to patrons.

The Attribute Gage R&R is just one of many MSA tools an organization may use to better understand their processes and verify the validity and utility of the data collected through their measurement systems. Such understanding contributes to the library organization's efforts to improve quality and respond to change through informed decision making.

Notes

1. Sarah Anne Murphy, "Leveraging Lean Six Sigma to Culture, Nurture, and Sustain Assessment and Change in the Academic Library Environment," *College & Research Libraries* 70, no. 3 (May 2009): 215–225.
2. American Library Association, Reference and User Services Association, *Guidelines for the Behavioral Performance of Reference and Information Service Providers* (Chicago: American Library Association). Available online at www.ala.org/ala/mgrps/divs/rusa/resources/guidelines/guide-linesbehavioral.cfm. [Accessed 3 November 2008].
3. American Library Association, Reference and User Services Association, *Definitions of Reference*, available online at www.ala.org/ala/mgrps/divs/rusa/resources/guidelines/definitionsreference.cfm [Accessed 3 November 2008]; Jo Bell Whitlatch, *Evaluating Reference Services: A Practical Guide* (Chicago: American Library Association, 2000); William A. Katz, *Introduction to Reference Work*, McGraw-Hill Series in Library Education (New York: McGraw-Hill, 1969); and American Library Association, *The Reference Assessment Manual* (Ann Arbor, Mich.: Pierian Press, 1995).
4. American Library Association, *Guidelines for the Behavioral Performance*.
5. Mathew L. Saxton and John V. Richardson, *Understanding Reference Transactions: Transforming an Art into a Science* (New York: Academic Press, 2002); Nahyun Kwon and Vicki L. Gregory, "The Effects of Librarians' Behavioral Performance on User Satisfaction in Chat Reference Services," *Reference & User Services Quarterly* 47, no. 2 (winter 2007): 137–48; Fu Zhuo, et al., "Applying RUSA Guidelines in the Analysis of Chat Reference Transcripts," *College & Undergraduate Libraries* 13, no. 1 (June 2006): 75–88.
6. William D. Mawby, *Make Your Destructive, Dynamic, and Attribute Measurement System Work for You* (Milwaukee, Wisc.: ASQ Quality Press, 2006).
7. Samuel E. Windsor, "Attribute Gage R&R," *Six Sigma Forum Magazine* 2, no. 4 (Aug. 2003): 23–28.
8. Murphy, "Leveraging Lean Six Sigma."
9. Michael Mueller, *Turning Judgment Calls into Reliable Data with Gage R&R*, available online at <http://europe.isixsigma.com/library/content/c041020b.asp> [Accessed 24 July 2008]; Joseph D. Conklin, "Measurement System Analysis for Attribute Measuring Processes," *Quality Progress* 39, no. 3 (2006): 50–53.
10. Windsor, "Attribute Gage R&R," 27.

Appendix A:

Quality Standards for Virtual Reference Policy

Policy: Quality Standards for Virtual Reference

This policy covers answers to questions OSU Libraries receives via QuestionPoint and instant messaging software, and is intended to support OSU Libraries efforts to provide consistent customer service to the Libraries' user population.

Policy Guidelines

I. Definitions for Classifying Question Type

A. Directional Question. "A directional transaction is an information contact that facilitates the logistical use of the library and that does not involve the knowledge, use, recommendations, interpretations, or instruction in the use of any information sources other than those that describe the library, such as schedules, floor plans, and handbooks."¹ This includes any question requiring a basic understanding of OSU library services and the OSU library organization to answer. It also includes questions requiring knowledge of how to operate equipment, including such things as printers, copiers, scanners, and computer hardware and software.

Examples:

- Circulation policies
- Fine disputes
- Returns, claims
- ILL, Article Express
- Reserves
- Referrals to technical departments within OSU libraries

- Referrals to other departments at OSU
- EZ-proxy/Off-campus sign-in issues
- RefWorks
- Questions about library programming
- Complaints
- Donation referrals

B. Reference Question. “A reference transaction is an information contact that involves the knowledge, use, recommendations, interpretation, or instruction in the use of one or more information sources by a member of the library staff. The term includes information and referral service. Information sources include (a) printed and no printed material; (b) machine-readable databases (including computer-assisted instruction); (c) the library’s own catalogs and other holdings records; (d) other libraries and institutions through communication or referral; and (e) persons both inside and outside the library. When a staff member uses information gained from previous use of information sources to answer a question, the transaction is reported as a reference transaction even if the source is not consulted again.”¹

1. Basic Reference Question. This includes any general questions that can be answered using the library’s own catalogs and other holdings records (i.e., OhioLINK, WorldCat); a standard library database, such as Academic Search Premier; a standard reference handbook, or annual; the OSU Libraries Web site; and referrals to subject specialists (when a patron specifically requests a referral to a subject specialist). Questions that require instruction on using the library catalog, such as how do I recall a book, should be coded Basic Reference.

2. Specialist Reference Question. This includes any question that requires specialist knowledge of a subject area and multiple sources to answer. (All questions that come in through the University Archives Ask-An-Archivist and Ohioline Web site should be assigned to this category.)

II. Criteria for a Quality Answer

Answers to questions received through the OSU Libraries Ask-A-Question e-mail and IM services should:

- Reflect The OSU Libraries Service Values;
- Conform to the *RUSA Guidelines for the Behavioral Performance of Reference and Information Service Providers*;
- Be free of library jargon;
- Acknowledge the patron or the patron’s question;
- Thank the patron for their question;
- Reflect ownership of the patron’s question or problem;
- Provide appropriate referrals;
- Not contain significant grammatical or spelling errors;
- Recognize patron affiliation;
- Provide permanent URLs for OSU Library Catalog records along with the call number and holding location for the item referenced;
- Provide URLs for Web sites referenced in the answer;
- Instruct users on how to utilize library and information resources (if appropriate); and
- Indicate that the individual answering the question comprehended the patron’s question.

In instances where the patron’s question is ambiguous or unclear, there should be 1) evidence that a qualifying question was asked by the individual answering the question; or 2) a statement that indicates how the question was interpreted, followed by an invitation for the patron to contact the individual answering the question if the answer doesn’t match the patron’s intended question.

III. Follow-up Language

The RUSA Guidelines note that the information service provider should “ask patrons if their questions have been completely answered” and “encourage patrons to return if they have further questions.”

IV. Service Denial

Service is considered denied when there is no evidence the question was answered via e-mail, phone, in person, or another communication mechanism. Service is also considered denied when a patron does not receive an answer because he or she is not a member of the OSU community or a resident of Ohio.

Procedure

I. Reference and Information Service Provider Responsibilities

- A. Answer virtual reference questions using the criteria for a quality answer.
- B. Include follow-up language in answers.
- C. Avoid service denial. Document reason for not answering a question within QuestionPoint.
- D. Arrange backup if you are going to miss your shift (IM) or be out of the office for longer than a 24-hour period Monday through Friday. Change the settings in your QuestionPoint account so that your e-mail notification is sent directly to your backup.
- E. For e-mail questions, answer question within 24 hours of receipt, Monday through Friday, if possible. Notify patron of delay if 24 hours is not possible.
- F. For IM questions, save transaction logs to IM server space monthly. When recording transactions in the Ask Database, write the IM client the question was received through in the comments section followed by a semicolon.
Example: meebo;

II. Coordinator of Research and Reference Responsibilities

- A. Coordinate monthly review of Virtual Reference transactions. Monitor the receipt and distribution of directional, basic, and specialist reference questions. Share results with CIPS department at least quarterly.
- B. Review policy annually by coordinating annual Attribute Gage R&R, to be executed with a rotating mix of faculty librarians and paraprofessional employees over summer quarter.

Reference

1. Association of Research Libraries. *ARL Statistics 2005–2006: A Compilation of Statistics from One Hundred and Twenty-Three Members of the Association of Research Libraries* (Washington, D.C.: Association of Research Libraries, 2008), 99.